

ABSTRACT OF THE DISCLOSURE

METHOD AND APPARATUS FOR DATA REDUNDANCY ELIMINATION AT THE BLOCK LEVEL

5

A redundancy elimination mechanism is provided, which applies aspects of duplicate block elimination and delta encoding at the block level. The redundancy elimination mechanism divides file objects into content-defined blocks or “chunks.” Identical chunks are suppressed. The redundancy elimination mechanism also performs
10 resemblance detection on remaining chunks to identify chunks with sufficient redundancy to benefit from delta encoding of individual chunks. Any remaining chunks that do not benefit from delta encoding are compressed. Resemblance detection is optimized by merging groups of fingerprints into super fingerprints. This merging can be constructed to ensure that if two objects have a single super fingerprint in common, they are
15 extremely likely to be substantially similar.